Linear regression using both temporally aggregated and temporally disaggregated data

Palm, F.C.; Nijman, T.E.

Published in:
Journal of Econometrics

Publication date:
1982

Journal of Econometrics 19 (1982) 333–343. North-Holland Publishing Company

LINEAR REGRESSION USING BOTH TEMPORALLY AGGREGATED AND TEMPORALLY DISAGGREGATED DATA

F.C. PALM and T.E. NIJMAN

Free University, 1007 MC Amsterdam, The Netherlands

Received June 1981, final version received March 1982

We consider the model introduced by Hsiao (1979) for analyzing data that are subject to different temporal aggregation. We show that maximum likelihood estimation of the parameters in the model is more complex than Hsiao suggests. Using a simple example we compare the large sample variance of the maximum likelihood estimator with that of the estimator proposed by Hsiao, of the generalized least squares (GLS) estimator and of the OLS estimator applied to aggregate data.

1. Introduction

We consider the model introduced by Hsiao (1979) for analyzing data that are subject to different temporal aggregation. He analyzes the problem of missing observations on a disaggregate level when a temporal aggregate, e.g. the total for a year, is observed. An attractive feature of his approach is that he relates the series for which some observations are missing to other explanatory variables, thereby taking account of the information in the observed temporal aggregate. His approach can be very useful when additional information in the form of e.g. causal relationships for the missing observations is available to the econometrician.

Using a maximum likelihood approach (ML), Hsiao states that the likelihood function can be factorized in two parts, so that the computational work required for ML estimation is substantially reduced. We shall show that the computation of ML estimates is more complex than Hsiao suggests. ML estimation requires joint estimation of all the parameters in the model.

In section 2, we set up the model and the density function for the observations. Section 3 is devoted to the estimation of the model. We derive the ML estimator and compare its large sample variance with that of the estimator proposed by Hsiao, the GLS estimator and the OLS estimator applied to aggregate data. In section 4, we briefly present some conclusions.

2. The model

In this note we limit ourselves to the simplest case of the model discussed by Hsiao (1979). Similar conclusions also hold for more general versions of the model.

Hsiao considers the following regression equation:

$$y_{ti}=\beta x_{ti}+u_{ti}, \qquad i=1,2, \quad t=1,2,\ldots,T, \tag{1}$$

where the u_{ti}'s are independent normal variates with mean zero and variance σ^2 and independent of x_{ti}'s. The subscript t denotes the year, i denotes whether the observation is made in the first or the second six-month period of year t. Assume that y_{ti} is observed in periods of six months whereas for x_{ti} only annual aggregates $\bar{x}_t=x_{t1}+x_{t2}$ are available. This is a problem of missing observations.

Hsiao extends his model by assuming that there exist observations on a series z_{ti} related to the variable x_{ti} through

$$x_{ti}=\alpha z_{ti}+v_{ti}, \tag{2}$$

where the v_{ti}'s are independently normally distributed with mean zero and variance σ_v^2 and independent of z_{ti}. For simplicity, he assumes that z_{ti} has mean zero and constant variance σ_z^2 and is independent of u_{ti}.

Along with several other authors, Gourieroux and Monfort (1981) consider a regression model in which the variable y_{ti} is explained by x_{ti} and z_{ti}, whereas for x_{ti}, for which some observations are missing, eq. (2) holds. Following an approach similar to that proposed by Anderson (1957), they reparametrize the joint distribution for y_{ti} and x_{ti} given z_{ti} as a product of the marginal distribution for y_{ti} given z_{ti} and the conditional distribution of x_{ti} given y_{ti} and z_{ti}. They show that this reparametrization which is one-to-one with the original parametrization of the joint distribution, provides an immediate solution for the ML estimator. However, because z_{ti} is not included in eq. (1), the parameters of the joint process for (y_{ti}, x_{ti}) are restricted in the present case. Therefore, the computational advantages of the reparametrization are lost. Dagenais (1973) analyzes the properties of regression coefficient estimates in a model similar to that used by Gourieroux and Monfort (1981), when the missing explanatory variables are approximated by linear functions of available independent variables.

Conditional on z_{ti}, the vector $(y_{ti}, x_{ti})'$ is normally distributed with mean $\mu=(\beta\alpha z_{ti}, \alpha z_{ti})'$ and covariance matrix

$$\Sigma=\begin{pmatrix}\sigma_v^2\beta^2+\sigma^2 & \sigma_v^2\beta \\ \sigma_v^2\beta & \sigma_v^2\end{pmatrix}.$$

The assumptions made in (1) and (2) imply one restriction on the parameters of the joint distribution of $(y_{ti}, x_{ti})'$, i.e., a restriction on the parameters of Σ.

The model is quite simple, but the likelihood approach proposed by Hsiao can also be applied when the model is extended

(a) for y_{ti}, x_{ti} and z_{ti} being vectors of random variables [see Hsiao (1979)],
(b) for more than 2 subperiods within a year,
(c) for the case where the distribution of z_{ti} is known [e.g. when z_{ti} is normally distributed, one should analyze the joint distribution of (y_{ti}, x_{ti}, z_{ti})],
(d) when z_{ti} is a non-stochastic variable such as a time trend allowing for a non-stationary mean of y_{ti} and x_{ti}.

After some transformations and integration with respect to x_{t1}, the joint distribution of $(\bar{y}_t, y_{t1}-y_{t2}, \bar{x}_t)$ given z_{t1} and z_{t2}, with $\bar{y}_t = y_{t1}+y_{t2}$, which is equivalent to the density of the observations, is obtained. It is a normal distribution with mean $m=[\alpha\beta\bar{z}_t, \alpha\beta(z_{t1}-z_{t2}), \alpha\bar{z}_t]'$ and covariance matrix

$$\Omega = \begin{bmatrix} 2\tilde{\sigma}^2 & 0 & 2\beta\sigma_v^2 \\ 0 & 2\tilde{\sigma}^2 & 0 \\ 2\beta\sigma_v^2 & 0 & 2\sigma_v^2 \end{bmatrix}, \tag{3}$$

where

$$\tilde{\sigma}^2 = (\sigma^2 + \beta^2\sigma_v^2).$$

For a sample of T years, the likelihood function can then be written as

$$L(\alpha, \sigma_v^2, \sigma^2, \beta \mid \text{data}) = (2\pi)^{-T}(2\sigma^2 2\tilde{\sigma}^2)^{-T/2}$$

$$\times \exp\left\{-\left(\frac{1}{4\sigma^2}\right)\sum_{t=1}^{T}(\bar{y}_t-\beta\bar{x}_t)^2 - \left(\frac{1}{4\tilde{\sigma}^2}\right)\sum_{t=1}^{T}[(y_{t1}-y_{t2})-\beta\alpha(z_{t1}-z_{t2})]^2\right\}$$

$$\times (2\pi)^{-T/2}(2\sigma_v^2)^{-T/2}\exp\left\{-\left(\frac{1}{4\sigma_v^2}\right)\sum_{t=1}^{T}(\bar{x}_t-\alpha\bar{z}_t)^2\right\}. \tag{4}$$

Hsiao (1979) erroneously assumes that only the last term of expression (4) depends on α and σ_v^2, so that the estimator for $\theta=(\alpha, \sigma_v^2, \sigma^2, \beta)'$ given on p. 247 is not the ML estimator. The ML estimator is obtained by maximizing the log-likelihood function with respect to θ.

3. Estimation of the model

The first-order conditions for a maximum of $\ln L$,

$$\frac{\partial \ln L}{\partial \alpha}=\frac{1}{2\sigma_v^2}\sum_{t=1}^{T}\bar{z}_t(\bar{x}_t-\alpha\bar{z}_t)$$

$$+\frac{1}{2\tilde{\sigma}^2}\sum_{t=1}^{T}\beta(z_{t1}-z_{t2})[y_{t1}-y_{t2}-\beta\alpha(z_{t1}-z_{t2})]=0,$$

$$\frac{\partial \ln L}{\partial \sigma_v^2}=\frac{-T\beta^2}{2\tilde{\sigma}^2}+\frac{\beta^2}{4\tilde{\sigma}^4}\sum_{t=1}^{T}[y_{t1}-y_{t2}-\beta\alpha(z_{t1}-z_{t2})]^2$$

$$-\frac{T}{2\sigma_v^2}+\frac{1}{4\sigma_v^4}\sum_{t=1}^{T}(\bar{x}_t-\alpha\bar{z}_t)^2=0,$$

$$\frac{\partial \ln L}{\partial \sigma^2}=\frac{-T}{2\sigma^2}-\frac{T}{2\tilde{\sigma}^2}+\frac{1}{4\sigma^4}\sum_{t=1}^{T}(\bar{y}_t-\beta\bar{x}_t)^2$$

$$+\frac{1}{4\tilde{\sigma}^4}\sum_{t=1}^{T}[y_{t1}-y_{t2}-\beta\alpha(z_{t1}-z_{t2})]^2=0,$$

$$\frac{\partial \ln L}{\partial \beta}=\frac{-T\beta\sigma_v^2}{\tilde{\sigma}^2}+\frac{1}{2\sigma^2}\sum_{t=1}^{T}\bar{x}_t(\bar{y}_t-\beta\bar{x}_t)$$

$$+\frac{\sigma_v^2\beta}{2\tilde{\sigma}^4}\sum_{t=1}^{T}[y_{t1}-y_{t2}-\beta\alpha(z_{t1}-z_{t2})]^2$$

$$+\frac{1}{2\tilde{\sigma}^2}\sum_{t=1}^{T}\alpha(z_{t1}-z_{t2})[y_{t1}-y_{t2}-\beta\alpha(z_{t1}-z_{t2})]=0, \tag{5}$$

are nonlinear in θ.

The system of equations in (5) can be solved iteratively. Two-step estimation, e.g. using the method of scoring and starting with an initial consistent estimate of θ, will achieve asymptotic efficiency

$$\hat{\theta}_2=\hat{\theta}_1-\left\{\mathrm{E}\left[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right]_{\theta=\hat{\theta}_1}\right\}^{-1}\left\{\frac{\partial \ln L}{\partial\theta}\right\}_{\theta=\hat{\theta}_1}, \tag{6}$$

where $\hat{\theta}_1$ is an initial consistent estimate of θ. The two-step estimator $\hat{\theta}_2$ has the same asymptotic distribution as the ML e 'imator, $\hat{\theta}_{\mathrm{ML}}$. The asymptotic covariance matrix is given by

$$\operatorname{var}(\sqrt{T}\hat{\theta}_2)=\operatorname{var}(\sqrt{T}\hat{\theta}_{\mathrm{ML}})=-T\left\{\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\theta\,\partial\theta'}\right]\right\}^{-1}$$

$$=\begin{bmatrix} \frac{\beta^2\tilde{\sigma}_z^2}{2\tilde{\sigma}^2}+\frac{\mathrm{E}\bar{z}_t^2}{2\sigma_v^2} & 0 & 0 & \frac{\alpha\beta\tilde{\sigma}_z^2}{2\tilde{\sigma}^2} \\ 0 & \frac{\beta^4}{2\tilde{\sigma}^4}+\frac{1}{2\sigma_v^4} & \frac{\beta^2}{2\tilde{\sigma}^4} & \frac{\beta^3\sigma_v^2}{\tilde{\sigma}^4} \\ 0 & \frac{\beta^2}{2\tilde{\sigma}^4} & \frac{1}{2\tilde{\sigma}^4}+\frac{1}{2\sigma^4} & \frac{\beta\sigma_v^2}{\tilde{\sigma}^4} \\ \frac{\alpha\beta\tilde{\sigma}_z^2}{2\tilde{\sigma}^2} & \frac{\beta^3\sigma_v^2}{\tilde{\sigma}^4} & \frac{\beta\sigma_v^2}{\tilde{\sigma}^4} & \frac{\mathrm{E}\bar{x}_t^2}{2\sigma^2}+\frac{\alpha^2\tilde{\sigma}_z^2}{2\tilde{\sigma}^2}+\frac{2\beta^2\sigma_v^4}{\tilde{\sigma}^4} \end{bmatrix}, \tag{7}$$

where

$$\tilde{\sigma}_z^2=\mathrm{E}(z_{t1}-z_{t2})^2,\quad \mathrm{E}\bar{z}_t^2=\mathrm{E}(z_{t1}+z_{t2})^2,\quad \mathrm{E}\bar{x}_t^2=\mathrm{E}(x_{t1}+x_{t2})^2.$$

Notice that the information matrix in (7) is not block diagonal, so that the elements in $\hat{\theta}_2$ are not independently distributed.

Estimation of one element in the vector affects the efficiency of the other elements. In order to achieve efficiency, the variances and the regression coefficients in the vector θ have to be estimated jointly. From (7), the asymptotic variance of $\hat{\beta}_{\mathrm{ML}}$ can be obtained,

$$\operatorname{var}(\sqrt{T}\hat{\beta}_{\mathrm{ML}})=\left\{\frac{\mathrm{E}\bar{x}_t^2}{2\sigma^2}+\frac{\alpha^2\tilde{\sigma}_z^2}{2\tilde{\sigma}^2}\left[1-\frac{\beta^2\tilde{\sigma}_z^2\sigma_v^2}{(\beta^2\tilde{\sigma}_z^2\sigma_v^2+\mathrm{E}\bar{z}_t^2\tilde{\sigma}^2)}\right]\right.$$
$$\left.+\frac{2\beta^2\sigma_v^4}{\tilde{\sigma}^4}\left[1-\frac{(\sigma^4+\beta^4\sigma_v^4)}{(\sigma^4+\beta^4\sigma_v^4+\tilde{\sigma}^4)}\right]\right\}^{-1}=q. \tag{8}$$

Obviously it is greater than the variance given by Hsiao (1979) in his expression (18), which is the large sample variance of $\sqrt{T}\ \hat{\beta}_{\mathrm{ML}}$, given α and σ_v^2,

$$\operatorname{var}(\sqrt{T}\,\hat{\beta}_{\mathrm{ML}}\,|\,\alpha,\sigma_v^2)=\left\{\frac{\mathrm{E}\bar{x}_t^2}{2\sigma^2}+\frac{\alpha^2\tilde{\sigma}_z^2}{2\tilde{\sigma}^2}+\frac{2\beta^2\sigma_v^4}{\tilde{\sigma}^4+\sigma^4}\right\}^{-1}=r. \tag{9}$$

Similarly, in expression (17) on p. 248, Hsiao gives the asymptotic variance of the generalized least squares estimator of β, $\sqrt{T}\,\hat{\beta}_{\mathrm{GLS}}$, given α and known or consistently estimated disturbance covariance matrix, applied to the two equations,

$$y_{ti}=\beta\hat{x}_{ti}+w_{ti}, \qquad i=1,2, \quad t=1,2,\ldots,T. \tag{10}$$

The explanatory variable in (10), $\hat{x}_{ti}=\alpha z_{ti}+\frac{1}{2}(\bar{x}_t-\alpha\bar{z}_t)$, is the minimum mean square error predictor of x_{ti}, given z_{t1}, z_{t2} and $\bar{x}_t$, and

$$w_{t1}=u_{t1}+\frac{\beta}{2}(v_{t1}-v_{t2}), \qquad w_{t2}=u_{t2}+\frac{\beta}{2}(v_{t2}-v_{t1}).$$

As stated by Hsiao, the large sample variance of the GLS estimator,

$$\mathrm{var}(\sqrt{T}\,\hat{\beta}_{\mathrm{GLS}}\,|\,\alpha)=\left\{\frac{\mathrm{E}\tilde{x}_t^2}{2\sigma^2}+\frac{\alpha^2\tilde{\sigma}_z^2}{2\bar{\sigma}^2}\right\}^{-1}=p, \tag{11}$$

is greater than that of the ML estimator in (9) when α and σ_v^2 are known. When α is not known but is estimated in a regression on the aggregated data $\bar{x}_t=\alpha\bar{z}_t+\bar{v}_t$, the asymptotic variance of the GLS estimator becomes

$$\mathrm{var}_{\hat{\delta}}(\sqrt{T}\,\hat{\beta}_{\mathrm{GLS}})=p+\frac{p^2\alpha^2\beta^2\sigma_v^2\tilde{\sigma}_z^4}{2\bar{\sigma}^4\mathrm{E}\bar{z}_t^2}, \tag{12}$$

which is different from that given in (11). Estimation of α increases the variance of $\sqrt{T}\,\hat{\beta}_{\mathrm{GLS}}$. This finding is at variance with a conclusion given by Dagenais (1973) for the GLS estimator when the missing observations are related to observed variables through parameters to be estimated. Notice that Dagenais assumes that at least $k+1$ values (k being the number of regressors in the original model) of the missing series are observed instead of the aggregates.

After some straightforward but tedious transformations a comparison of (12) with (8) shows that the GLS estimator is less efficient than the ML estimator when α has to be estimated.

When $\delta=(\alpha,\sigma_v^2)'$ is unknown, one can substitute a consistent estimate $\hat{\delta}$ for it and compute the two-step estimator of $\gamma=(\sigma^2,\beta)'$ presented by Hsiao (H) in his expression (13) on p. 247,

$$\hat{\gamma}_2=\hat{\gamma}_1-\left\{\mathrm{E}\left[\frac{\partial^2\ln L_1}{\partial\gamma\partial\gamma'}\bigg|_{\substack{\gamma=\hat{\gamma}_1\\ \delta=\hat{\delta}}}\right]\right\}^{-1}\frac{\partial\ln L_1}{\partial\gamma}\bigg|_{\substack{\gamma=\hat{\gamma}_1\\ \delta=\hat{\delta}}}, \tag{13}$$

where $\hat{\gamma}_i$ denotes the ith step estimator of γ, L_1 is the density of $(y_{t1}-y_{t2},\bar{y}_t)$, $t=1,\ldots,T$, given in (4). In the appendix we briefly indicate how to derive the variance of the estimator in (13), when α and σ_v^2 are obtained by OLS applied to the aggregate data

$$\operatorname{var}_{\hat{\alpha},\hat{\sigma}_v^2}(\sqrt{T}\hat{\beta}_{\mathrm{H}}) = r + r^2\left\{\frac{\alpha^2\beta^2\tilde{\sigma}_z^4\sigma_v^2}{2\tilde{\sigma}^4\mathrm{E}\bar{z}_t^2} + \frac{2\sigma_v^8\beta^6}{(\tilde{\sigma}^4+\sigma^4)^2}\right\}. \tag{14}$$

Again, expression (14) can be shown after some transformations to be greater than the variance of the ML estimator in (8).

Of course, when α and σ_v^2 are known, the variance of Hsiao's estimator is equal to the variance of the ML estimator, r, given in (9).

Finally, if we apply OLS to the aggregated data $\bar{y}_t = \beta\bar{x}_t + \bar{u}_t$, the asymptotic variance of $\sqrt{T}\hat{\beta}_{\mathrm{OLS}}$ is equal to $2\sigma^2/\mathrm{E}\bar{x}_t^2$, both for α known and α unknown.

The relationships between the variances are summarized in table 1.

Table 1
Asymptotic variances for the estimators of β.

α, σ_v^2	OLS, aggregated data		GLS		Hsiao		ML
Unknown	$2\sigma^2/\mathrm{E}\bar{x}_t^2$		$p+\frac{p^2\alpha^2\beta^2\tilde{\sigma}_z^4\sigma_v^2}{2\tilde{\sigma}^4\mathrm{E}\bar{z}_t^2}$	$\geqq$	$r+r^2\left\{\frac{\alpha^2\beta^2\tilde{\sigma}_z^4\sigma_v^2}{2\tilde{\sigma}^4\mathrm{E}\bar{z}_t^2}+\frac{2\sigma_v^8\beta^6}{(\tilde{\sigma}^4+\sigma^4)^2}\right\}$	$\geqq$	q
	$=$		$\geqq$		$\geqq$		$\geqq$
Known	$2\sigma^2/\mathrm{E}\bar{x}_t^2$	$\geqq$	p	$\geqq$	r	$=$	r

In order to give the reader an impression of the loss of efficiency involved when using a method other than ML, we have computed the ratio of the asymptotic variance of the different estimators with respect to the variance of the ML estimator.

The results are presented in table 2.

In the first four columns of table 2 we report the values of the parameters in the model, i.e., the variance of z_{ti}, the correlation between z_{t1} and z_{t2}, α and β, and σ_v^2 and σ^2.

In column 5, we report the theoretical coefficient of determination of the regression equations for x_t and y_t, respectively.

We have used the coefficients of determination and the correlation between z_{t1} and z_{t2} to select the values of the parameters for which results are reported in table 2. Notice also that the relative efficiency of the estimators is completely determined by the two R^2's and the correlation between z_{t1} and z_{t2}.

Table 2
Relative efficiency of the ML estimator compared with alternative estimators for β, measured by the ratio of large sample variances.[a]

var(Z)	corr(Z)	α, β	σ_v^2, σ^2	R_x^2, R_y^2		OLS	GLS	Hsiao
91.837	0.800	0.700	5.000	0.900	U	1.072	1.018	1.001
		0.800	3.556	0.900	K	1.078	1.022	1.000
91.837	0.800	0.700	5.000	0.900	U	1.101	1.003	1.000
		0.800	48.000	0.400	K	1.102	1.003	1.000
6.803	0.800	0.700	5.000	0.400	U	1.078	1.069	1.027
		0.800	0.593	0.900	K	1.126	1.116	1.000
6.803	0.800	0.700	5.000	0.400	U	1.159	1.112	1.000
		0.800	8.000	0.400	K	1.166	1.118	1.000
91.837	0.400	0.700	5.000	0.900	U	1.198	1.026	1.005
		0.800	3.556	0.900	K	1.238	1.024	1.000
91.837	0.400	0.700	5.000	0.900	U	1.367	1.003	1.000
		0.800	48.000	0.400	K	1.377	1.003	1.000
6.803	0.400	0.700	5.000	0.400	U	1.102	1.080	1.030
		0.800	0.593	0.900	K	1.165	1.129	1.000
6.803	0.400	0.700	5.000	0.400	U	1.264	1.119	1.001
		0.800	8.000	0.400	K	1.288	1.122	1.000
91.837	0	0.700	5.000	0.900	U	1.355	1.059	1.032
		0.800	3.556	0.900	K	1.513	1.026	1.000
91.837	0	0.700	5.000	0.900	U	1.800	1.004	1.001
		0.800	48.000	0.400	K	1.850	1.003	1.000
6.803	0	0.700	5.000	0.400	U	1.125	1.111	1.046
		0.800	0.593	0.900	K	1.217	1.145	1.000
6.803	0	0.700	5.000	0.400	U	1.376	1.138	1.010
		0.800	8.000	0.400	K	1.448	1.126	1.000
91.837	−0.400	0.700	5.000	0.900	U	1.544	1.185	1.142
		0.800	3.556	0.900	K	2.097	1.030	1.000
91.837	−0.400	0.700	5.000	0.900	U	2.621	1.008	1.004
		0.800	48.000	0.400	K	2.855	1.003	1.000
6.803	−0.400	0.700	5.000	0.400	U	1.144	1.228	1.117
		0.800	0.593	0.900	K	1.288	1.167	1.000
6.803	−0.400	0.700	5.000	0.400	U	1.469	1.209	1.053
		0.800	8.000	0.400	K	1.669	1.131	1.000
91.837	−0.800	0.700	5.000	0.900	U	1.697	1.766	1.666
		0.800	3.556	0.900	K	4.185	1.034	1.000
91.837	−0.800	0.700	5.000	0.900	U	4.494	1.032	1.027
		0.800	48.000	0.400	K	6.446	1.003	1.000
6.803	−0.800	0.700	5.000	0.400	U	1.153	2.056	1.669
		0.800	0.593	0.900	K	1.393	1.195	1.000
6.803	−0.800	0.700	5.000	0.400	U	1.438	1.725	1.428
		0.800	8.000	0.400	K	1.995	1.136	1.000

[a]The symbols U and K are used to indicate that the parameters α and σ_v^2 are assumed to be respectively unknown and known, when estimating β.

Although the estimator proposed by Hsiao is not an ML estimator when α and σ_v^2 have to be estimated, its efficiency is almost as high as that of the ML estimator, except when the correlation between z_{t1} and z_{t2} becomes negative and large in absolute value.

A negative correlation between z_{t1} and z_{t2} can arise when z_{ti} represents e.g. expenditures, which can be shifted from the first half to the second half of the year and vice versa, leaving the total for the year unaffected.

For the last but one model in table 2, OLS is more efficient than the estimator proposed by Hsiao (α and σ_v^2 being unknown).

Notice also that in four cases for α unknown the variance of the OLS estimator applied to aggregate data is smaller than that of the GLS estimator. A look at the results in table 2 shows that the loss of efficiency when using GLS or OLS instead of ML can sometimes be substantial.

4. Concluding remarks

The treatment of the problem of missing observations and the estimation of the unknown parameters proposed by Hsiao is very promising. However, the computations involved are more complex than suggested by Hsiao (1979) and Dagenais (1973), if the aim is to achieve efficiency. All estimators are consistent.

The estimator proposed by Hsiao is still fairly precise compared to the ML estimator when α is unknown, except when z_{t1} and z_{t2} are strongly negatively correlated. Full asymptotic efficiency requires joint estimation of all the parameters in θ.

The GLS and the ML estimators of β usually have different large sample properties, both with α known and α unknown. These conclusions also hold true in more general models.

Finally, as stated by Hsiao, the efficient estimation of the unknown parameters and the best predictions (in the sense of minimum mean square error) of time series by related series are inseparable. In order to predict the unknown series x_{t1}, one might prefer to use all the related series, i.e., to use

$$\begin{aligned} \mathrm{E}(x_{t1} \mid z_{t1}, z_{t2}, y_{t1}, y_{t2}, \bar{x}_t) = {} & \alpha z_{t1} + \frac{\beta\sigma_v^2}{2\bar{\sigma}^2}(y_{t1} - \beta\alpha z_{t1}) \\ & - \frac{\beta\sigma_v^2}{2\bar{\sigma}^2}(y_{t2} - \beta\alpha z_{t2}) + \tfrac{1}{2}(\bar{x}_t - \alpha\bar{z}_t) \end{aligned}$$

as a predictor instead of using

$$\mathrm{E}(x_{t1} \mid z_{t1}, z_{t2}, \bar{x}_t) = \alpha z_{t1} + \tfrac{1}{2}(\bar{x}_t - \alpha\bar{z}_t).$$

Appendix

We shall give the main steps of the derivation of the large sample variance for the estimator proposed by Hsiao [see expression (13) given above]

$$\hat{\gamma}_2 = \hat{\gamma}_1 - A \frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\hat{\gamma}_1\\ \delta=\hat{\delta}}} \tag{A.1}$$

where

$$\delta = (\alpha, \sigma_v^2)', \qquad \gamma = (\sigma^2, \beta)', \qquad A = \mathrm{E}\left[\frac{\partial^2 \ln L_1}{\partial \gamma \partial \gamma'}\bigg|_{\substack{\gamma=\hat{\gamma}_1\\ \delta=\hat{\delta}}}\right]^{-1}.$$

A Taylor series expansion of the first-order derivatives in (A.1) around the estimate $\hat{\delta}$ and the true value of γ, γ_0, gives

$$\frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\hat{\gamma}_1\\ \delta=\hat{\delta}}} = \frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\gamma_0\\ \delta=\hat{\delta}}} + \frac{\partial^2 \ln L_1}{\partial \gamma \partial \gamma'}\bigg|_{\substack{\gamma=\bar{\gamma}\\ \delta=\hat{\delta}}} (\hat{\gamma}_1 - \gamma_0), \tag{A.2}$$

with

$$|\bar{\gamma} - \gamma_0| \leqq |\hat{\gamma}_1 - \gamma_0|.$$

After substitution of (A.2) into (A.1) we get

$$\sqrt{T}(\hat{\gamma}_2 - \gamma_0) = \left(I - A \frac{\partial^2 \ln L_1}{\partial \gamma \partial \gamma'}\bigg|_{\substack{\gamma=\bar{\gamma}\\ \delta=\hat{\delta}}}\right) \sqrt{T}(\hat{\gamma}_1 - \gamma_0) - \sqrt{T}\, A \frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\gamma_0\\ \delta=\hat{\delta}}} \tag{A.3}$$

The first factor of the first r.h.s. term of (A.3) has zero probability limit, so that this term converges to zero in probability, provided $\hat{\gamma}_1$ has a well behaved limiting distribution [see e.g. Dhrymes and Taylor (1976, pp. 370–371)].

The second r.h.s. term in (A.3) can be written as

$$-\sqrt{T} A \frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\gamma_0\\ \delta=\hat{\delta}}} = -\sqrt{T} A \frac{\partial \ln L_1}{\partial \gamma}\bigg|_{\substack{\gamma=\gamma_0\\ \delta=\delta_0}}$$

$$-\sqrt{T} A \frac{\partial^2 \ln L_1}{\partial \gamma \partial \delta'}\bigg|_{\substack{\gamma=\gamma_0\\ \delta=\bar{\delta}}} (\hat{\delta} - \delta_0), \tag{A.4}$$

with

$$|\bar{\delta}-\delta_0| \leqq |\hat{\delta}-\delta_0|,$$

being the true value of δ.

The first r.h.s. term of (A.4) is the ML estimator of γ given the true value of δ. Therefore it is asymptotically efficient with covariance matrix TA and it is not correlated with $\sqrt{T}(\hat{\delta}-\delta_0)$ in the second r.h.s. term of (A.4).

The asymptotic variance of $\hat{\gamma}_2$ is equal to the sum of TA and the variance of the second r.h.s. term in (A.4), which is a linear transformation of $\sqrt{T}(\hat{\delta}-\delta_0)$. It can be computed once an estimator $\hat{\delta}$ has been chosen.

References

Anderson, T.W., 1957, Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, Journal of the American Statistical Association 52, 200–203.

Dagenais, M.G., 1973, The use of incomplete observations in multiple regression analysis: A generalized least squares approach, Journal of Econometrics 1, 317–328.

Dhrymes, Ph.J. and J.B. Taylor, 1976, On an efficient estimator for dynamic simultaneous equations models with autoregressive errors, International Economic Review 17, 362–376.

Gourieroux, Ch. and A. Monfort, 1981, On the problem of missing data in linear models, Review of Economic Studies 48, 579–586.

Hsiao, Cheng, 1979, Linear regression using both temporally aggregated and temporally disaggregated data, Journal of Econometrics 10, 243–252.